

1. Assignment NMST 539

Frantisek Helebrand

February 2022

For this assignment, we used the dataset *ushealth* which is available in the *SMSdata* package. In total, we have 50 observations (from each US state) and 13 variables. In this assignment, we will focus on the number of cardiovascular deaths and regions, where the patients lived. The regions are Northeast, Midwest, South, West. In our data, we have 9 states in the Northeast, 12 states in the Midwest, 16 states in the South and 13 in the West region.

In this assignment, we will be examining the dependence between the number of cardiovascular deaths and the region i. e. if we can say that a patient from one region is more likely to die from cardiovascular illness than the patient from another region.

The figure 1 shows the distribution of cardiovascular deaths in each region. We can notice that the distribution of cardiovascular deaths with respect to the West region is slightly different. The expected value and variance is greater than other regions.

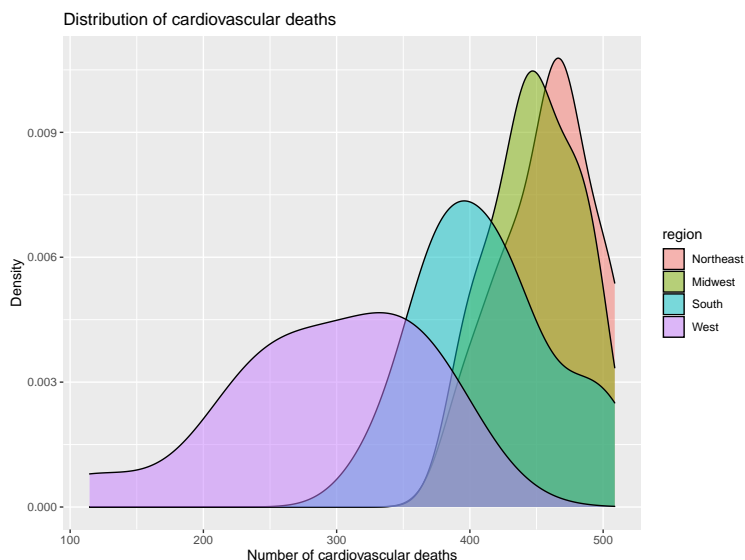


Figure 1

The independence can be seen in the figure 2 which shows the boxplots of cardiovascular deaths with respect to the region. We can see that boxplots for the Northeast, Midwest are similar but the boxplots for the west and south are different. From here we can see that number of cardiovascular deaths and regions are not independent. However, this is not a rigorous statistical test and further examination is needed.

